Uncertainty quantification and visualization for functional random variables

S. Nanty
Université Joseph Fourier, Grenoble, France

Affiliation: CEA, DEN, DER, F-13108 Saint Paul-lez-Durance, France

Email: simon.nanty@cea.fr – **URL**: http://www-ljk.imag.fr/

Master: Université Jean Monnet, Saint-Etienne, France

Ph.D. (2012-2015): Université Joseph Fourier, Grenoble, France

Supervisor(s): C. Prieur (Laboratoire Jean Kuntzman) and C. Helbert (Ecole Centrale Lyon) and A. Marrel (CEA) and N. Pérot (CEA)

Abstract:

In the framework of industrial risk assessment studies, the reliability of a component is evaluated during accidental conditions. A numerical model provides the thermal-hydraulic (T-H) loading of this component subjected to highly hypothetical conditions. Then, a numerical model for the mechanical analysis of components and structures (called hereafter the T-M code), taking as input the T-H loading, calculates the breaking strength of the component and the thermo-mechanical actual applied load. From these two elements, a safety criterion is deduced.

Under the hypothesis of accidental conditions, the component behavior depends on many uncertain parameters, related to the initial plant conditions or to the safety system characteristics. These uncertain parameters are input variables of the T-H code. They can be of various types: scalar, functional, categorical... It can be important to assess how these uncertainties can affect the code forecasts. To deal with all these uncertainties, computer experiment methodologies like uncertainty propagation and sensitivity analysis are useful. As the T-H code is much more time expensive than the T-M code, the uncertainties on the results of the T-H code which are of functional type, are directly characterized, so that uncertainty propagation and sensitivity analysis can be applied on the T-M code.

The uncertainty characterization of functional input variables has already been investigated by a few authors. A common way to study functional variables is to decompose them on a functional basis. Collin et al. [1] decompose the functional variable under study thanks to Functional Principal Component Analysis (FPCA), developed by Ramsay and Silverman [4]. As they consider that the functional random variable is a Gaussian process, the coefficients of the FPCA basis functions are independent and follow a centered and standardized Gaussian distribution. However, the Gaussian process hypothesis is quite restrictive in practice. Hyndman and Shang [3] propose a method close to the previous one. The functional variable is first decomposed on a FPCA basis. The joint probability density function of the coefficients on this basis is then estimated thanks to a kernel density estimator (Rosenblatt [5]). However, the kernel density estimation is inefficient in high dimension, so that the number of functions in the FPCA basis must be small to apply this method in practice.

In the present work, the problem under consideration is different from the previously studied ones in the sense that the functional random variables to be characterized are dependent upon one another and are linked to a scalar (or vectorial) variable, called hereafter a covariable. The considered covariable is here the output of the second code. This covariable can be, for instance, the output of a computer code which takes as inputs the functional random variables. The main objective of this work is thus to provide a new methodology to characterize the uncertainties associated to dependent functional variables linked to a covariable. These functional variables are discretized in practice. As in the two presented methodologies, the proposed characterization process is composed of two parts. First, the dimension of the problem is reduced by decomposing the functional random variables on a functional basis. In order to take into account the dependence between the functional random variables, the decomposition is done simultaneously on all the studied functional random variables. This means that the decomposition is

done on a vector of functional random variables instead of a functional random variable. We propose a simultaneous version of classical Partial Least Squares (Wold [6]), denoted SPLS. This SPLS method enables to take into account the link between the functional random variables and the covariable. The functional random variables are approximated by their coefficients on the SPLS basis. The problem becomes then multivariate instead of functional. The second step of the characterization procedure consists in estimating the joint probability density function of the basis coefficients. The distribution of these coefficients is modeled by a Gaussian mixture, whose parameters are estimated by the Expectation-Maximization algorithm (Dempster et al. [2]). Thanks to this modeling, the probability density function can be estimated in higher dimension than with the kernel density estimation, which is usually inefficient above dimension 6, so that a higher number of SPLS basis functions can be selected. This methodology gives a statistical characterization of the uncertainties on the studied functional random variables. New realizations of these random variables can also be simulated: coefficients are sampled from the estimated Gaussian mixture model, and the corresponding functions are then built by multiplying the new coefficients with the SPLS basis functions. The presented methodology has been tested and validated on a numerical example with two dependent functional variables and on a nuclear application with three dependent functional variables.

References

[1] F. Collin, J. Goffart, T. Mara, and L. Denis-Vidal. Uasa of complex models: Coping with dynamic and static inputs. Sensitivity Analysis of Model Ouputs Conference, 2013.

[2] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 1–38, 1977.

[3] R. J. Hyndman and H. L. Shang. Rainbow plots, bagplots, and boxplots for functional data. *Journal of Computational and Graphical Statistics*, 19(1):29–45, 2010.

[4] J. O. Ramsay and B. W. Silverman. *Functional Data Analysis*. Springer Series in Statistics. Springer, 2nd edition edition, 2005.

[5] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, pages 832–837, 1956.

[6] H. Wold. Path models with latent variables: The nipals approach. *Quantitative sociology: international perspectives on mathematical and statistical modeling*, 1975.

Short biography – After a master's degree in applied mathematics in Saint-Etienne, Simon Nanty began his PhD in october 2012 with Grenoble university and the comissariat à l'énergie atomique. This PhD thesis is funded by the CEA, and its objective is to quantify the uncertainties associated to a computer code with functional inputs and scalar output.